DIS 5D

Thursday, July 19, 2018 11:50 AM

Topic: Variance. Joint Distribution (cont.)

Variance "deviation"

- $Var(X) = E((X-E(X))^2) = E(X^2) - (E(X))^2$
- A closer look at $E(X^2)$...

if X is the number of something, it's a common strategy to break up X into indicators (give each "object" one indicator).

if $X = \sum_{i=1}^{n} X_i$ where X_i's are indicators, then

$$E(X^2) = n\,E(X_i^2) + n(n-1)\,E(X_iX_j)$$
$$= n\,\mathbb{P}(X_i=1) + n(n-1)\,\underbrace{\mathbb{P}(X_i=1, X_j=1)}_{X_i,\,X_j \text{ might not be independent!}}$$

joint distribution

- Distribution of one RV ...

either this of it as a one-dimensional table:

x	x_0	x_1	...
$X=x$			

$\mathbb{P}(X=x_0)$ $\mathbb{P}(X=x_1)$

or a function with one variable:

$P_X(x) = f(x) = \mathbb{P}(X=x)$

plug in x, it'll give you the probability of the event "$X=x$"

- Joint Distribution of 2 RVs ...

either think of it as a 2-dimensional table:

	x_0	x_1	...	x_i	...
y_0					
y_1					
⋮					
y_j					
⋮					

$\mathbb{P}(X=x_1, Y=y_0)$ $\quad$ $\mathbb{P}(X=x_i, Y=y_j)$

Fill in the entire table like this

or a function that has two variables

$P_{X,Y}(x,y) = f(x,y) = \mathbb{P}(X=x, Y=y)$

plug in x and y, it'll give you the probability of the event "$X=x, Y=y$"

- We can learn a lot from this 2-D table!

Specifically, conditional distribution and marginal distribution

→ Conditional Distribution

a fancy way of saying "looking at one row / one column".

e.g.

Y \ X	0	1	2
0	0.1	0.05	0.15
1	0.2	0.1	0.1
2	0.15	0.1	0.05

This is a valid joint distribution since probability sums up to 1.

$\boxed{\text{Q: What is the distribution of } X \mid Y=1\text{ ?}}$

This means, we want to know the distribution of X at this row

Y \ X	0	1	2
0	0.1	0.05	0.15
1	0.2	0.1	0.1
2	0.15	0.1	0.05

Let's directly copy this distribution into a 1-D table:

x	0	1	2
$\mathbb{P}(X=x \mid Y=1)$	0.2	0.1	0.1

This does not seem correct the probability is not 1!

Let's scale each term so it sums up to 1:

x	0	1	2
$\mathbb{P}(X=x \mid Y=1)$	$\frac{0.2}{0.4}$	$\frac{0.1}{0.4}$	$\frac{0.1}{0.4}$

Looks pretty good now ☺

Let's see what we did above:

$$\mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$$

This exactly how we would calculate the conditional probability of two events.

→ Marginal Distribution

a fancy way of saying $\mathbb{P}(X=x)$ or $\mathbb{P}(Y=y)$

You can use joint distribution to recover distribution of each RV!

e.g.

Y \ X	0	1	2
0	0.1	0.05	0.15
1	0.2	0.1	0.1
2	0.15	0.1	0.05

Q: What's the distribution of X?

Let's calculate $\mathbb{P}(X=0)$ first.

When $X=0$, Y can be 0, 1, or 2.

That is, $\mathbb{P}(X=0) = \mathbb{P}(X=0, Y=0) + \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=0, Y=2)$

$= 0.1 + 0.2 + 0.15$

$= 0.45$

We write this probability in the margin like this:

Y \ X	0	1	2
0	0.1	0.05	0.15
1	0.2	0.1	0.1
2	0.15	0.1	0.05
	0.45		

Keep doing the same thing to get the distribution of X:

x	0	1	2
$\mathbb{P}(X=x)$	0.45	0.25	0.3

Let's generalize this:

$\mathbb{P}(X=x) = \sum_{\forall y} \mathbb{P}(X=x, Y=y)$ marginal distribution of X

$\mathbb{P}(Y=y) = \sum_{\forall x} \mathbb{P}(X=x, Y=y)$ marginal distribution of Y